The long-term differential achievement effects of school socioeconomic composition in primary education: A propensity score matching approach

British Journal of Educational Psychology (2016), 86, 501–525

The British Psychological Society

www.wileyonlinelibrary.com

The long-term differential achievement effects of school socioeconomic composition in primary education: A propensity score matching approach

Barbara Belfi[1,2]*, Carla Haelermans[3] and Bieke De Fraine[1]

[1]Department of Educational Sciences, Center for Educational Effectiveness and Evaluation, Katholieke Universiteit Leuven, Belgium
[2]Research Centre for Education and the Labour Market, Maastricht University, The Netherlands
[3]Top Institute of Evidence Based Education Research, Maastricht University, The Netherlands

Background. The effects of school socio-economic composition on student achievement growth trajectories have been a hot topic of discussion among politicians around the world for many years. However, the bulk of research investigating school socio-economic composition effects has been limited in important ways.

Aims. In an attempt to overcome the flaws in earlier research on school socio-economic composition effects, this study used data from a large sample, followed students throughout primary education, addressed selection bias problems, identified the grade(s) in which school socio-economic composition mattered the most, and studied the differential effects of school socio-economic composition by individual socio-economic status (SES).

Sample. In a longitudinal design with seven occasions of data collection, the authors drew on a sample of $N = 3{,}619$ students (age at T1 about 5 years, age at T7 about 12 years) from 151 primary schools in Flanders (the northern part of Belgium).

Method. Students in low-, medium-, high-, and mixed-SES schools were matched using propensity scores. To compare students' achievement growth trajectories in the different school compositions, multilevel regression modelling with repeated measurements was applied.

Results. The results showed that students had more positive achievement growth in high-SES as compared to low-SES and mixed-SES schools. In two of the three comparisons, students in mixed-SES schools showed the lowest math development. The negative effects of mixed-SES schools on math achievement growth were the strongest for high-SES students.

Conclusions. Our findings contribute to the ongoing discussion on the effects of school socio-economic composition on student achievement growth.

*Correspondence should be addressed to Barbara Belfi, Research Centre for Education and the Labour Market, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands (email: b.belfi@maastrichtuniversity.nl).

DOI:10.1111/bjep.12120

Ever since the famous 1966 Coleman report concluded that 'the social composition of the student body is more highly related to achievement, independent of the student's own social background, than is any school factor' (Coleman *et al.*, 1966, p. 325), school socio-economic composition has been a popular research subject in the field of educational effectiveness research (Hattie, 2002). It has also been a topic of interest for politicians around the world who fear that large proportions of socio-economically disadvantaged students in schools will have a detrimental effect on the educational trajectories of all students attending such schools. In particular, in Flanders, the Dutch-speaking part of Belgium, where socio-economic school segregation is very high compared to other Western countries, disparities in school socio-economic composition have been a major cause for concern (Jacobs, Rea, & Teney, 2009; OECD, 2010). As such, school socio-economic desegregation has often been proposed as a policy lever to address socio-economic disparities in education.

Over the past few decades, a multitude of research has focussed on socio-economic composition effects in primary schools. However, this research has been limited in a number of ways. In particular, three issues have been largely neglected in school compositional research. First, it is problematic that most studies investigating primary school socio-economic composition effects have been cross-sectional or have used a pre/post-design (Agirdag, Van Houtte, & Van Avermaet, 2012; Battistich, Solomon, Kim, Watson, & Schaps, 1995; Driessen, 2002; Dumay & Dupriez, 2008; Opdenakker, Van Damme, De Fraine, Van Landeghem, & Onghena, 2002; Peetsma, Van der Veen, Koopman, & Van Schooten, 2006; Strand, 1997; Van der Slik, Driessen, & De Bot, 2006; Willms, 2010). Research investigating the long-term effects of primary school socio-economic composition is scarce (Belfi *et al.*, 2014; Guldemond & Bosker, 2009; Verhaeghe, Van Damme, & Knipprath, 2011). Moreover, this research only seldom identifies the grade(s) in which the achievement gaps between students attending schools with different socio-economic compositions decrease or increase. However, knowing in which grades school socio-economic composition matters the most is important as it enables both practitioners and policymakers to more efficiently tackle achievement gaps associated with school socio-economic composition.

Second, there is a lack of research examining whether school socio-economic composition effects hold equally for different groups of students. Research using a pre/post-design has already indicated that although a high percentage of low-socio-economic status (SES) students in schools negatively affected high-SES students' achievement, it positively affected low-SES students' achievement (Peetsma *et al.*, 2006). Thus, it is possible that low-SES schools are ultimately not as detrimental to low-SES students' long-term educational development as is often thought. Such information on differential effects of school composition would be valuable for policymakers proposing school desegregation as a strategy to raise low-SES students' educational performance. However, to date, no lines of research have investigated this.

A third and final hiatus in school socio-economic composition research involves the methodology used. Most scholars investigating school composition effects use multilevel regression models in which they control a number of student- and school-related covariates. However, this analytic approach does not optimally controls for selection bias problems, as regression models extrapolate over portions of the covariate space where there are no data (Gelman & Hill, 2007). As such, regression models may over- or underestimate effects by making comparisons in sections where there is no clear counterfactual for either group. As a result, we are unable to decide whether a widening achievement gap between schools with different socio-economic composition is a result of the socio-economic composition of the school or whether it is simply the result of previously existing differences among students (Retelsdorf, Becker, Köller, & Möller, 2012). One methodology that is more suitable to identify comparable individuals, address selection bias, and estimate causal effects is propensity score matching (Rosenbaum & Rubin, 1983). Unlike regression analyses, this method does not rely on partialing out explanatory covariates from the outcome measures. Instead, it directly controls the assignment process for individuals and models this process instead of the outcome (Schafer & Kang, 2008). To our knowledge, primary school socio-economic composition effects have not been explored by means of propensity score matching so far.

Present study

In an attempt to overcome the aforementioned flaws in the research literature on primary school socio-economic composition, this study uses data from a large sample, follows students throughout primary education, addresses the problem of selection bias using propensity score matching, identifies the grade(s) in which school socio-economic composition matters the most, and studies the differential effects of school socio-economic composition by individual SES. Based on these aspects, the following research questions were formulated:

1. What are the long-term effects of primary school socio-economic composition on students' mathematics achievement growth?
2. Are the long-term effects of primary school socio-economic composition different for students with different individual SES levels?

Method

Sample

The data used in this study stem from the SiBO project (i.e., the Dutch acronym for School Trajectories in Primary Education; Maes, Ghesquiere, Onghena, & Van Damme, 2002), a large-scale longitudinal research project designed to describe and explain interindividual differences in students' developmental trajectories throughout primary education in Flanders. The SiBO project involved a nationally representative random sample of 120 schools which was extended with an oversampling of 31 schools with a high percentage of low-SES students. The SiBO data set includes observations from a total of N = 3,619 students in 151 schools. Because students were tested at different time points during the period from 2002 to 2009, the total sample size differed for each measurement occasion (see Table 1). Students who repeated a grade or changed school were included in the analyses until the moment of grade repetition or school changing took place, as these students are part of the Flemish school reality.[1]

Main study variables

Both student-level and school-level variables were studied. All variables were standardized to have a mean of zero and a standard deviation of one to express the parameter estimates in effect size units. Below, the main study variables are described.

[1] *Preliminary analyses showed that the percentage of students that were retained or changed school did not vary by school SES type.*

Table 1. Number of schools and students per school socio-economic status (SES) type and per measurement occasion

	Low-SES schools		Medium-SES schools		High-SES schools		Mixed-SES schools	
	N schools	*N* students	*N* schools	*N* students	*N* schools	*N* students	*N* schools	*N* students
Grade 1	10	167	105	2,644	14	226	34	582
Grade 2	10	146	101	2,394	8	188	30	516
Grade 3	10	123	98	2,142	7	181	30	463
Grade 4	9	99	96	2,004	8	169	30	418
Grade 5	10	88	99	1,892	7	159	29	387
Grade 6	10	80	96	1,765	9	148	30	340

Student SES

Student SES was constructed using five indicators: (1) educational level of the father, (2) educational level of the mother, (3) occupation of the father, (4) occupation of the mother, and (5) family income. Information on these indicators was drawn from the parent questionnaire that was administered in kindergarten (school year 2002–2003). Next, one factor score was calculated using confirmatory factor analysis. Estimation of the measurement model indicated a good fit, $\chi^2(27) = 1811.49$, $p < .01$; TLI = .96; CFI = .97; RMSEA = .01 (Hu & Bentler, 1995). The completely standardized factor loadings ranged between .75 and .85 and were significant ($p < .001$). This variable's internal consistency was good (Cronbach's $\alpha = .82$). To test whether student SES remained stable over time and could be regarded as a time-invariant variable, we calculated the Pearson correlation between two SES measures: (1) the SES measure that is used in this study and which was constructed on the basis of information drawn from the parent questionnaire that was administered in kindergarten (school year 2002–2003) and (2) a similar SES measure based on the parent questionnaire administered in Grade 4 (school year 2006–2007). These two SES measures were found to be highly correlated ($r = .89$, $p < .001$). Furthermore, in Table 2, the Pearson correlations between the separate SES indicators of school year 2002–2003 and 2006–2007 are shown. The correlations range between .69 (occupation of the father) and .89 (educational level of the mother) and are significant ($p < .001$), further indicating that there was low social

Table 2. Pearson correlations between the indicators of the socio-economic status (SES) variable of school 2002–2003 and the SES variable of 2006–2007

School year 0203 – School year 0607	
1. Educational level of the father	.87***
2. Educational level of the mother	.89***
3. Occupation of the father	.69***
4. Occupation of the mother	.71***
5. Family income	.75***

Note. ***$p < .001$.

mobility over the 4 years and that student SES could be regarded to be time-invariant indeed.

School socio-economic composition

School socio-economic composition was defined by the percentage of low-, medium-, or high-SES students in the school in the school year 2002–2003. In Flanders, schools with a high share of low-SES students are commonly named 'concentration schools', which is a pejorative term. In terms of absolute thresholds, a school is called a 'concentration school' when it includes at least 50% low-SES students (Dors, Karsten, Ledoux, & Steen, 1991; Driessen, 2002). Hence, to test the possible differential effects of school socio-economic composition, four meaningful categories were distinguished: (1) low-SES schools (schools with ≥50% low-SES students), (2) medium-SES schools (schools with ≥50% medium-SES students), (3) high-SES schools (schools ≥50% high-SES students), and (4) mixed-SES schools (all schools that do not fit the qualifications for the other three school types). Low-SES students were defined as students with a SES of 1 *SD* or more below the average (i.e., mean); high-SES students were defined as students with a SES of 1 *SD* or more above the average; and medium-SES students were defined as students with an average SES (a SES between −1 *SD* and +1 *SD*). The descriptive statistics of the resulting four school SES types are shown in Table 3. To confirm that school SES was a time-invariant variable, we calculated the Pearson correlation between the mean school SES in the school year 2002–2003 and the mean school SES in the school year 2006–2007. As this correlation was high ($r = .95, p < .001$), we concluded that school SES was stable over time.

Gender

Student gender was registered by means of a dummy variable. Girls were coded 0, and boys were coded 1.

Age

Student age was registered at the beginning of kindergarten (in months).

Math achievement

Student math achievement was examined through curriculum-based math achievement tests that were especially designed for the SiBO study (Cortois, Van Droogenbroeck, Verachtert, & Van Damme, 2011). In grades 1–6, math achievement tests were administered to all students at the end of each school year. Each math achievement test consisted of 50–80 items, covering the following domains in mathematics: number sense,

Table 3. Descriptives of the average socio-economic status (SES) of the four school SES types

	N	*M*	*SD*	Min	Max
Low-SES schools	167	−1.10	0.52	−2.19	0.79
Medium-SES schools	2,644	0.03	0.83	−2.36	2.07
High-SES schools	226	0.59	0.86	−1.87	2.07
Mixed-SES schools	582	−0.14	1.02	−2.19	2.07

number procedures, measurement, geometry, and applied math problem-solving. Cronbach's alpha coefficients ranged between .88 and .94 across all waves, indicating high internal consistency. Importantly, after administration, students' scores on all test items were vertically linked based on common items using a three-parameter IRT model with Bayes model estimates (Verhaeghe, 2010). Based on this equation, students' raw mathematic scores could be converted to math IRT θ scores with a common metric. These IRT θ scores were used in this study, allowing us to compare students' math performance across grades and across time.

School ethnic composition

School ethnic composition was operationalized as the percentage of non-Western students in each school. Following Driessen (2002), Belfi *et al*. (2014) and Belfi, Gielen, De Fraine, Verschueren, and Meredith (2015), we used the birthplace of a student's mother to assess student ethnicity. Students whose mother was born in a Western country (i.e., Western Europe, the United States, and Australia) were coded 0, and students whose mother was born in a non-Western country were coded 1. To confirm that school ethnic composition was a time-invariant variable, we calculated the Pearson correlation between the percentage of non-Western in the school year 2002–2003 and the percentage of non-Western in the school year 2006–2007. As this correlation was high (r = .94, p < .001), we concluded that school ethnic composition was stable over time.

School achievement composition

School achievement composition was operationalized as the school-level aggregation of student achievement on a pre-mathematic test for kindergarteners Cortois *et al.*, 2011; Cronbach's α: .93) at the end of the school year 2002–2003. To confirm that school size was a time-invariant variable, we calculated the Pearson correlation between aggregated mathematic scores in the school year 2002–2003 and the aggregated mathematic scores the school year 2006–2007. As this correlation was high (r = .72, p < .001), we concluded that school achievement composition was stable over time.

School size

School size was measured as the total number of students enrolled in the school at the beginning of the school year 2002–2003. To confirm that school size was a time-invariant variable, we calculated the Pearson correlation between the school size in the school year 2002–2003 and the school year 2006–2007. As this correlation was high (r = .92, p < .001), we concluded that school size was stable over time.

Missing data

Missing data present a considerable practical problem in all longitudinal research, including this study. Data were missing because students were absent, because some schools decided not to participate at a certain time point, or because teachers or parents did not return the questionnaires. In this study, there was an average of 23.7% missing data across all points of measurement. Under the missing at random condition (MAR) condition, the missingness mechanism is ignorable. However, the MAR distribution is a strong assumption and it is impossible to test whether this condition is satisfied, except by gathering data from non-respondents (Pinxten, De Fraine, Van Damme, & D'Haenens, 2010; Schafer & Graham, 2002). Nonetheless, even if the data are not MAR, multiple imputation is thought to produce more accurate estimates of population parameters than would be obtained if listwise deletions were used (Schafer & Graham, 2002). Therefore, to take advantage of all data provided by the participants, multiple imputation was applied. Missing values were imputed using the STATA implementation (ICE, Royston, 2004) of the MICE program (Van Buuren & Groothuis-Oudshoorn, 2011). All variables were used as predictors for the other variables, with exception of the outcome variable, which was not used to impute pre-treatment variables (Langenskiöld, 2005). In total, five complete data sets were generated, with the missing values being replaced by plausible values. After data imputation, matching was conducted for each of the imputed data sets separately.[2]

Analyses

Propensity score matching

Propensity score matching has become an increasingly popular approach to control for potential selection bias (Rosenbaum & Rubin, 1983). Through propensity score matching, we can estimate the average treatment effect (ATT) on the treated rather than the average treatment effect (ATE; e.g., Imai, King, & Stuart, 2008). In this study, the ATE characterizes the average effect of attending a certain school SES type on all students, even those who would never attend such a school SES type. However, by applying propensity score matching, we are able to analyse the effect of attending a certain school SES type only on those students who would typically attend such a school SES type. The propensity score represents the conditional probability of receiving a treatment given observable characteristics selecting individuals into treatment (Rosenbaum & Rubin, 1983). In the present research, the propensity score is the conditional probability of being enrolled in one of the four school SES types (i.e., low-, medium-, high-, or mixed-SES school), rather than in one of the three other school SES types, given a set of 17 observed covariates. Prior empirical research was used to identify background characteristics that increase a child's probability of being enrolled in either one of the four school SES types. This set of background characteristics was comprised of demographics (gender, age), social background indicators (educational level of mother and father), ethnic background (birth country of student, mother, father, maternal and paternal grandparents; language spoken at home), prior achievement (language and math achievement score at the end of the final kindergarten grade; information on whether the student was advised to follow special education at the end of the final kindergarten year), and family situation indicators (information regarding whether the student was adopted and lived together with both his/her parents). Information on the covariates was gathered from several achievement tests and the parent questionnaire administered in the final year of kindergarten. Prior to estimating the propensity scores, simple *t*-test statistics were used to determine whether mean differences in background characteristics of students enrolled in the four school SES types were statistically significant. Table 4 summarizes the findings regarding the differences in background characteristics between students attending low-SES, medium-SES, high-SES, and mixed-SES schools. As the *t*-test results in Table 4 show, students

[2] *To ensure that the multiple imputation approach was appropriate, we compared the analyses with and without the imputed indicators in the equations. In both types of analysis, all parameters showed very similar results. Based on this 'sensitivity' analysis, it is likely that we properly handled the missing data challenge using multiple imputation.*

Table 4. Findings for the central covariates for the four school socio-economic status (SES) types at T1 before Kernel matching

Covariates	Low-SES schools (1)		Medium-SES schools (2)		High-SES schools (3)		Mixed-SES schools (4)		*t*-test results					
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	1–2	1–3	1–4	2–3	2–4	3–4
Student gender	−0.08	1.00	0.03	1.00	0.13	1.00	−0.01	1.00	−1.39	−2.01	−0.75	−1.35	0.99	1.78
Student age at the beginning of kindergarten (in months)	0.06	1.33	−0.05	0.92	−0.01	0.97	0.10	1.15	1.50	0.63	−0.37	−0.63	−3.43**	−1.29
Student language achievement score at the end of kindergarten	−0.69	1.05	0.16	0.92	0.17	0.92	−0.11	1.01	−11.52***	−8.54***	−6.47***	−0.06	6.35***	3.57***
Student math achievement score at the end of kindergarten	−0.89	0.83	0.21	0.91	0.16	0.94	−0.11	1.01	−15.28***	−11.50***	−9.16***	0.83	7.67***	3.54***
Student lives together with both parents	0.29	1.20	−0.08	0.92	−0.10	0.90	0.15	1.11	4.92***	3.73***	1.46	0.38	−5.12***	−3.01**
Student is adopted	0.30	1.20	−0.09	0.92	−0.09	0.91	0.14	1.11	5.20***	3.73***	1.67	0.13	−5.13***	−2.81**
Student was advised to follow special education at the end of kindergarten	0.21	1.62	−0.07	0.68	−0.05	0.76	0.08	1.28	4.56***	2.17**	1.09	−0.28	−3.98***	−1.50
Educational level mother	−0.65	0.78	0.12	0.94	0.46	1.15	−0.21	1.05	−10.39***	−10.74	−4.99***	−5.00***	7.64***	7.91***
Educational level father	−0.46	0.89	0.07	0.96	0.48	1.06	−0.21	1.07	−6.91***	−9.27***	−2.75**	−6.16***	6.17***	8.23***
Language spoken at home	0.34	1.25	−0.12	0.87	0.17	1.14	0.15	1.13	6.34***	1.40	1.84	−4.57***	−6.31***	0.19

Continued

Table 4. (*Continued*)

Covariates	Low-SES schools (1)		Medium-SES schools (2)		High-SES schools (3)		Mixed-SES schools (4)		*t*-test results					
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	1–2	1–3	1–4	2–3	2–4	3–4
Country of birth student	0.71	1.33	−0.21	0.76	−0.21	0.76	0.34	1.22	14.23***	8.63***	3.32**	0.02	−14.00***	−6.35***
Country of birth mother	0.15	1.33	−0.06	0.83	0.11	1.26	0.13	1.30	3.05**	0.29	0.16	−2.86**	−4.52***	−0.20
Country of birth father	0.41	1.26	−0.11	0.88	−0.05	0.95	0.18	1.14	7.24***	4.14***	2.29*	−1.01	−6.78***	−2.66**
Country of birth maternal grandmother	0.41	1.25	−0.13	0.87	−0.06	0.94	0.24	1.17	7.56***	4.27***	1.68	−1.12	−8.59***	−3.43**
Country of birth paternal grandmother	0.40	1.21	−0.13	0.88	−0.06	0.95	0.24	1.15	7.34***	4.17***	1.50	−1.20	−8.78***	−3.51**
Country of birth maternal grandfather	0.40	1.24	−0.13	0.87	−0.07	0.94	0.26	1.17	7.48***	4.27***	1.42	−1.08	−9.09***	−3.70***
Country of birth paternal grandfather	0.44	1.24	−0.15	0.86	−0.01	1.00	0.25	1.17	8.25***	3.92***	1.77	−2.30*	−9.45***	−2.95**

Note. The *t*-test results refer to the comparisons between each of the six school SES types: low-SES (1), medium-SES (2), high-SES (3), and mixed-SES schools (4).
*$p < .05$, **$p < .01$, ***$p < .001$.

enrolled in the four school SES types differed significantly on most covariates. To mitigate the impact of these covariates, propensity score matching was applied.

Propensity score matching is a two-stage process. Stage 1 involves *estimating the propensity score*. In this study, a logistic regression model predicting the probability for a student to be enrolled in either a low-, medium-, high-, or mixed-SES school rather than in the other three school SES types was estimated according to the 17 observed covariates. This predicted odds value became the propensity score on which students enrolled in a given school SES type were matched to similar students enrolled in one of the other three school SES types. Stage 2 entails the *actual matching* of the treated group (in our case: students enrolled in one of the four school SES types) to the non-treated groups (in our case: students enrolled in one of the other three school SES types) in such a way that the students in the four school SES types are equivalent on the observed covariates included in the propensity score. Students in the four school SES type groups were matched using the kernel matching approach, by means of the psmatch2 routine in Stata (Leuven & Sianesi, 2003). Kernel matching uses the whole sample of control individuals at each time but weights the observations depending on the proximity to the target individual in the treatment group. In kernel matching, the bandwidth parameter can be varied thus changing the degree to which the discrete density distribution of the propensity score is smoothed. Higher bandwidth parameters facilitate the inclusion of individuals (i.e., their consideration as 'similar') and thus increase efficiency but risk higher bias (Harder, Stuart, & Anthony, 2010). In this study, the lowest, most restrictive bandwidth parameter was used (i.e., bw = .02) rather than the average (i.e., bw = .06) or a less restrictive value (i.e., bw = .12).

Estimation of the ATTs

To estimate the effect of attending a low-, medium-, high-, or mixed-SES school on students' mathematics growth trajectories, the growth trajectories of students attending a certain school SES type were compared with the growth trajectories of similar students in one of the other three school SES types by means of multivariate linear regression analyses. As such, six comparative analyses were performed in which students in low-SES school were compared with similar students in medium-SES schools (comparison 1), high-SES schools (comparison 2), and mixed-SES schools (comparison 3); students in medium-SES schools were compared with similar students in high-SES schools (comparison 4) and mixed-SES schools (comparison 5); and students in high-SES schools were compared with similar students in mixed-SES schools (comparison 6).

Student achievement growth was analysed by fitting a three-level, repeated-measurements model using Stata 14 (StataCorp, 2015). A major advantage of conducting multilevel analyses with repeated measures is that these analyses do not require the same number of measurement occasions for each individual subject (Snijders & Bosker, 2012). Hence, students only had to participate in at least one measurement occasion to be included in the analyses. In each three-level model, students were nested within schools and measurement occasions were nested within students. As it is very well possible that student achievement growth differs from grade to grade, we opted for a saturated piecewise growth curve modelling approach, in which the total amount of achievement growth is divided into separate 'growth phases' based on a difference score between two measurement occasions. Because we only had two measurement occasions per 'growth phase', it was impossible to estimate the variance between two measurement occasions within pupils. Growth trajectories were thus estimated. The intercept in the growth

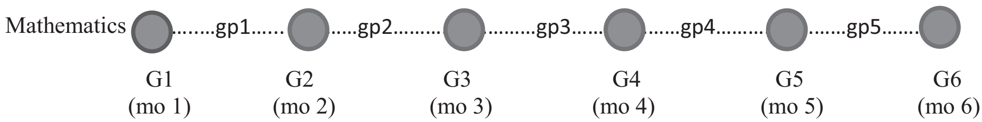

Figure 1. Overview of the measurement occasions and the growth phases of mathematics. *Note*. gp = growth phase; G = Grade; mo = measurement occasion.

model denotes the average achievement score at the end of Grade 1, and the following regression coefficients each represent a 'growth phase' for each set of following grades. As such, the model consisted of five growth phases (see Figure 1). The coefficients for the intercept and the growth phases were set randomly at both the student and school levels, thereby allowing for the estimations of achievement differences between students and between schools. To estimate the impact of school socio-economic composition, interaction terms were included in the models. Figure 1 shows the measurement occasions and the accompanying growth phases.

As indicated by Willms (2006, 2010) and Manski (2000), it is vital to take into account selection effects when examining compositional effects. A selection effect occurs when students with similar characteristics tend to be clustered in similar schools. To ensure that possible effects of school socio-economic composition are only due to the socio-economic make-up of the school and not to other school composition types closely related to this type of composition, several control variables should be included in the model. In this study, we include school achievement composition, school ethnic composition, and school size as control variables, because previous research has indicated that school socio-economic composition is associated with poorer school achievement (Belfi *et al.*, 2015), a higher proportion of ethnic minority students (Belfi *et al.*, 2014, 2015), and a larger student population (Van Maele & Van Houtte, 2011). As these different school composition types were only correlated moderately (see Table 5), it was concluded including them simultaneously in the model would not cause multicollinearity problems.

Estimation of the ATTs for students with a different SES

In a third step of the analyses, an interaction term between individual student SES and school SES was included in the models to investigate whether school socio-economic composition has differential effects for student with different individual SES levels.

Results

Propensity score matching results

The success of the matching procedure was tested by the check of balance between the treatment and control groups (whether matching homogenized the groups successfully) and the inspection of the area of common support (whether comparable individuals are available for the whole sample or only in some parts of the propensity score distributions). As an indication of a satisfying balance, Caliendo and Kopenig (2005) have indicated bias after matching should be below 3–5%. Table 6 shows the results of the kernel matching procedure. As can be seen, balance was only found in 2.94% of the terms, thereby indicating that the matching was successful in terms of creating homogeneous groups.

Table 5. Pearson correlations for the school composition types

	1.	2.	3.	4.
1. School socio-economic composition	–			
2. School ethnic composition	−.67***	–		
3. School achievement composition	.70***	−.70***	–	
4. School size	.08***	−.16***	.16***	–

Note. ***p < .001.

When comparing students within the area of common support that were used for matching with those that were not used for matching, the bias between all groups also was low (<5%). Thus, the results of the subsequent analyses may be taken for representative for students within the area of common support.[3]

Analysis of the ATTs of school socio-economic composition

After homogenous groups were created by means of propensity score matching, comparative analyses were performed in which the mathematic growth (Grade 1–Grade 6) of students in low-SES school was compared with similar students in medium-SES schools (comparison 1), high-SES schools (comparison 2), and mixed-SES schools (comparison 3); students in medium-SES schools were compared with similar students in high-SES schools (comparison 4) and mixed-SES schools (comparison 5); and students in high-SES schools were compared with similar students in mixed-SES schools (comparison 6). Table 7 summarizes the findings of these comparative analyses and presents the standardized regression coefficients as calculated by Cohen's d, while controlling the data's multilevel nature and other types of school composition.

With regard to comparisons between students in low-SES and medium-SES schools (comparison 1), students in low-SES and mixed-SES schools (comparison 3), and students in medium- and high-SES schools (comparison 4), no differences in math performance on either the first measurement occasion or achievement growth were found. However, among the other three comparisons, significant differences in student achievement growth were found indeed. First, it was found that while students in high-SES schools showed lower achievement at the first measurement occasion as compared to similar students in low-SES schools (d = −6.87, p < .001), they showed significantly more growth during the first (d = 10.72, p < .001) and second growth phase (d = 5.22, p < .05), thereby surpassing their peers in low-SES schools by the end of Grade 6 (see Table 8; comparison 3). Similarly, students in high-SES schools also showed more achievement growth than comparable peers in mixed-SES schools (comparison 6). Particularly, in the third growth phase, students in mixed-SES schools earned 2.31 points (p < .05) less as compared to their akin peers in high-SES schools. Finally, students in mixed-SES schools were found to achieve 2.97 points (p < .001) more than comparable matches in medium-SES schools at the first measurement occasion (comparison 5). However, during growth phase 1 (d = −2.48, p < .05) and growth phase 3 (d = −1.55,

[3] *As advised by Harder* et al. *(2010), we performed different matching techniques to choose the one that yielded the best balance of covariates and propensity scores (i.e., nearest neighbour matching with and without replacement using the ratios 1:1 and 1:5 as well kernel matching with the bandwidth parameters .06 and .12). The kernel matching approach with the bandwidth parameter .02 proved to yield the best results.*

Table 6. Findings for the central covariates for the four school socio-economic status (SES) types at T1 after Kernel matching

	Low-SES schools (1)		Medium-SES schools (2)		High-SES schools (3)		Mixed-SES schools (4)		t-test results					
Covariates	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	1–2	1–3	1–4	2–3	2–4	3–4
Student gender	−0.06	1.00	−0.02	1.00	−0.08	1.00	−0.07	1.00	−1.21	−0.94	−0.89	1.77	0.23	0.00
Student age at the beginning of kindergarten (in months)	0.05	1.33	0.05	1.01	0.06	1.33	0.06	1.33	0.04	−0.72	−0.16	0.92	0.06	0.12
Student language achievement score at the end of kindergarten	−0.67	1.06	−0.64	0.96	−0.69	1.05	−0.69	1.06	−0.74	−0.09	−1.13	0.25	1.62	0.24
Student math achievement score at the end of kindergarten	−0.86	0.82	−0.85	0.80	−0.89	0.83	−0.89	0.84	−0.43	0.00	−0.05	2.75**	1.13	0.51
Student lives together with both parents	0.26	1.18	0.25	1.17	0.29	1.20	0.30	1.20	0.30	−0.66	−0.24	−1.61	0.98	0.05
Student is adopted	0.26	1.18	0.25	1.17	0.30	1.20	0.29	1.20	0.19	−0.58	−0.39	−1.72	0.92	0.16
Student was advised to follow special education at the end of kindergarten	0.12	1.40	0.07	1.25	0.21	1.62	0.22	1.63	1.03	0.18	0.26	−1.12	−0.15	−0.33
Educational level mother	−0.63	0.77	−0.63	0.85	−0.65	0.78	−0.64	0.77	0.00	0.22	−0.27	1.01	0.44	−0.16
Educational level father	−0.44	0.89	−0.45	0.87	−0.46	0.89	−0.46	0.90	0.47	−0.11	−0.19	0.56	1.79	−0.33
Language spoken at home	0.32	1.25	0.33	1.3	0.34	1.25	0.34	1.26	−0.22	1.32	−0.53	0.25	−2.27*	0.46
Country of birth student	0.66	1.32	0.63	1.32	0.71	1.33	0.70	1.33	0.57	0.42	−1.04	−1.27	0.06	−0.26
Country of birth mother	0.13	1.29	0.16	1.34	0.15	1.33	0.12	1.27	−0.61	0.48	−0.96	−0.87	0.66	−0.25
Country of birth father	0.39	1.25	0.35	1.23	0.40	1.27	0.42	1.27	0.80	0.04	0.40	−2.22*	0.27	−0.08
Country of birth maternal grandmother	0.39	1.24	0.39	1.24	0.41	1.25	0.42	1.25	0.08	0.77	0.20	−1.86	−0.15	−0.52
Country of birth paternal grandmother	0.38	1.21	0.37	1.24	0.40	1.21	0.40	1.22	0.13	−0.01	0.37	−0.69	0.47	−0.04
Country of birth maternal grandfather	0.38	1.23	0.38	1.23	0.40	1.24	0.41	1.24	−0.06	0.86	0.16	−1.52	−0.16	−0.37
Country of birth paternal grandfather	0.42	1.24	0.41	1.23	0.44	1.25	0.44	1.25	0.17	−0.55	0.35	−1.75	0.12	0.10

Note. The *t*-test results refer to the comparisons between each of the six school SES types: low-SES (1), medium-SES (2), high-SES (3), and mixed-SES schools (4).
$^{*}p < .05$, $^{**}p < .01$.

Table 7. Results of math growth of students in low-, medium-, high-, and mixed- socio-economic status (SES) schools after Kernel matching

	Comparison 1		Comparison 2		Comparison 3		Comparison 4		Comparison 5		Comparison 6	
	Low-SES school (a) versus medium-SES school (b)		Low-SES school (a) versus high-SES school (b)		Low-SES school (a) versus mixed-SES school (b)		Medium-SES school (a) versus high-SES school (b)		Medium-SES school (a) versus mixed-SES school (b)		High-SES school (a) versus mixed-SES school (b)	
	b(d)	*SE*	*b(d)*	*SE*	*b(d)*	*SE*	*b(d)*	*SE*	*b(d)*	*SE*	*b(d)*	*SE*
Intercept[a]	66.22***	0.50	67.05***	0.50	66.67***	0.65	67.71***	0.79	67.73***	0.60	68.28***	0.72
School (b)[b]	−0.55	1.58	−7.62***	2.86	1.96	1.68	−0.92	1.32	3.49**	1.13	3.23	1.94
Growth phase 1	11.85***	0.49	12.61***	0.83	11.64***	0.76	11.48***	1.28	11.16***	0.48	11.53***	0.66
Growth phase 2	20.68***	0.61	21.72***	1.06	19.57***	1.12	21.03***	1.07	19.64***	0.96	19.87***	1.24
Growth phase 3	28.68***	0.58	31.48***	0.83	27.20***	0.86	29.12***	1.23	26.04***	0.77	26.56***	0.93
Growth phase 4	34.30***	0.70	37.18***	1.16	32.17***	1.09	34.99***	1.12	30.78***	0.88	31.75***	0.97
Growth phase 5	38.59***	0.53	40.72***	1.21	36.47***	0.94	38.68***	1.25	34.38***	1.05	35.73***	1.16
Growth phase 1*school (b)	2.39	1.45	10.72***	2.72	−0.94	1.78	1.58	1.32	−2.48*	1.17	−3.48	2.10
Growth phase 2*school (b)	1.91	1.23	5.22*	2.34	−0.50	1.68	2.42	1.39	−2.05	1.12	−3.61	1.97
Growth phase 3*school (b)	−0.16	1.14	0.91	1.30	−2.35	1.56	1.76	0.94	−1.55*	0.74	−2.31*	1.04
Growth phase 4*school (b)	0.61	1.27	−0.59	1.39	0.16	1.25	1.04	0.81	−0.98	0.72	−1.47	1.04
Growth phase 5*school (b)	−1.38	1.11	−0.56	0.62	−1.73	1.13	−0.04	0.42	−0.11	0.69	−0.40	0.86
School ethnic composition	−0.41	0.62	−0.59	0.50	0.20	0.59	−0.49	0.60	−0.01	0.50	−0.63	0.81
Growth phase 1*ethnic composition	−1.21	0.64	−1.04	0.89	−1.12	0.80	−0.44	0.77	−0.76	0.44	0.45	0.91
Growth phase 2*ethnic composition	−0.20	1.19	0.20	1.56	−0.96	0.98	−0.32	0.75	−0.46	0.60	0.49	1.16
Growth phase 3*ethnic composition	0.25	0.91	0.34	1.20	−0.65	0.87	0.61	0.73	−0.58	0.65	−0.19	1.19
Growth phase 4*ethnic composition	1.10	1.06	0.96	1.49	−0.10	1.10	0.09	0.70	0.03	0.66	0.04	1.11
Growth phase 5*ethnic composition	1.02	0.90	0.95	1.46	−0.12	0.95	−0.42	0.55	−0.16	0.63	0.73	1.12
School math composition	2.40***	0.46	2.72***	0.39	2.67***	0.49	3.03***	0.40	2.96***	0.33	2.56***	0.55
Growth phase 1*math composition	−1.62	0.73	−3.31***	0.81	−1.30	0.89	−0.94	0.67	−0.34	0.34	0.34	0.99
Growth phase 2*math composition	−1.57	0.96	−3.99***	1.15	−2.06*	1.09	−1.26	0.61	−0.38	0.53	−0.15	1.02
Growth phase 3*math composition	−0.66	0.72	−3.61***	0.88	−0.75	0.70	−1.34	0.59	−0.54	0.39	−0.86	1.03

Continued

Table 7. (*Continued*)

	Comparison 1		Comparison 2		Comparison 3		Comparison 4		Comparison 5		Comparison 6	
	Low-SES school (a) versus medium-SES school (b)		Low-SES school (a) versus high-SES school (b)		Low-SES school (a) versus mixed-SES school (b)		Medium-SES school (a) versus high-SES school (b)		Medium-SES school (a) versus mixed-SES school (b)		High-SES school (a) versus mixed-SES school (b)	
	b(*d*)	*SE*	*b*(*d*)	*SE*	*b*(*d*)	*SE*	*b*(*d*)	*SE*	*b*(*d*)	*SE*	*b*(*d*)	*SE*
Growth phase 4*math composition	−0.70	0.87	−2.67***	1.73	−0.98	0.93	−0.94	0.53	−0.36	0.45	−0.06	0.87
Growth phase 5*math composition	−1.11	0.59	−3.30***	1.19	−1.00	0.82	−1.12	0.46	−0.53	0.49	0.36	0.97
School size	0.04	0.48	3.19***	0.42	−0.46	0.69	0.20	0.49	−0.12	0.34	0.65	0.78
Growth phase 1*school size	0.40	0.47	−3.01*	1.26	1.31	1.73	−0.88	0.70	0.40	0.33	−0.60	0.96
Growth phase 2*school size	0.28	0.65	−4.34***	1.05	0.78	0.70	−1.10	0.61	−0.03	0.46	−1.01	1.26
Growth phase 3*school size	−0.41	0.66	−4.02**	1.19	0.92	0.92	−0.57	0.51	0.38	0.41	0.26	1.09
Growth phase 4*school size	0.24	0.76	−2.42	1.35	1.13	0.98	−0.11	0.37	0.25	0.41	0.37	0.98
Growth phase 5*school size	0.24	0.58	−4.11**	1.26	1.48	1.34	−0.46	0.45	0.50	0.46	0.32	1.31

Note. [a]The intercept refers the math score of a student in school (a) of each comparison at the end of Grade 1.
[b]School (b) refers to the value-added of school (b) in each comparison.
***$p < .001$, **$p < .01$, *$p < .05$.

$p < .05$), they grow less, thereby falling behind on their medium-SES school peers by the end of Grade 6 (see Table 8).

Notably, in all three comparisons in which significant differences in achievement growth were found, school socio-economic composition appeared to have only an impact on the achievement growth gained in the first 3 years of primary education. When we compare the student math scores of the four school types at the end of primary education (Grade 6) across all six comparisons, we find that, while there are large differences across the different comparisons, on average, students in high-SES schools showed the highest amount of achievement growth and students in mixed-SES schools the lowest (see Table 8).[4]

While other school composition variables in the model are not the primary concern of this article, it is worth mentioning that in none of the comparisons, ethnic school composition was found to have an effect on mathematic achievement and development, once school socio-economic composition was controlled for. School achievement composition was in all the comparisons associated with higher achievement at the first measurement occasion. Remarkably, school achievement was associated with less achievement growth in comparisons 2 and 3, once school socio-economic composition was controlled for. Finally, school size was only related to less achievement growth in comparison 2.

Analysis of the differential ATTs of school socio-economic composition by individual student SES

In a final step of the analyses, we tested whether the ATEs were different according to one's individual SES. The results of these differential analyses are presented in Table 9. Only in comparisons 3 and 6, statistically significant differential effects by individual SES were found. First, students with one 1 *SD* increase in their individual SES level, grew 2.22 points less during the third growth phase and in a mixed-SES than in a low-SES school, as compared to students with an average SES (comparison 3). Likewise, students with one 1 *SD* increase in their individual SES level, grew 1.62 points less during the first growth phase and in a mixed-SES than in a high-SES school, as compared to students with an average SES. From these results, it becomes clear that that the higher the individual SES of a student, the stronger the negative effects of mixed-SES schools on the math achievement growth were (as compared to low- and high-SES schools).

Discussion

In the current study, we sought to examine the effect of school socio-economic composition on children's mathematical development during a period that is believed to be most strongly influenced by socio-economic compositional aspects of the environment: the primary school period. This is because primary school entrance introduces most children to a more diverse social system for the first time in their lives. Until that point, children spent most of their time within the narrow family and community environments, which are often socially homogeneous (Benner & Crosnoe, 2011). The results showed

[4] *We also applied a doubly robustness check by controlling for all of the covariates that we used as predictors in the matching procedure in the regression analyses. This technique has been found to make the estimation of the effects less sensitive to model misspecification (Ho, Imai, King, & Stuart, 2007). However, as we did not find any parameter differences by applying this robustness check, we chose to report the results of the simpler model, without the covariates as control variables.*

Table 8. Total math score at the end of Grade 6 of comparable students in the four school types across all comparisons

	Comparison 1 Low-SES school versus medium-SES school	Comparison 2 Low-SES school versus high-SES school	Comparison 3 Low-SES school (a) versus mixed-SES school (b)	Comparison 4 Medium-SES school (a) versus high-SES school (b)	Comparison 5 Medium-SES school (a) versus mixed-SES school (b)	Comparison 6 High-SES school (a) versus mixed-SES school (b)	Average total math score across all comparisons
Low-SES school	200.32	210.76	193.72	x	x	x	201.60
Medium-SES school	200.32	x	x	203.01	189.73	x	197.69
High-SES school	x	219.08	x	203.01	x	193.72	205.27
Mixed-SES school	x	x	193.72	x	189.19	191.41	191.44

Note. SES = socio-economic status.
The total math score at the end of Grade 6 was calculated by adding the statistically significant coefficients ($p < .05$) of the intercept and growth phases from Table 6.

Table 9. Results of math growth of students with different individual socio-economic status (SES) in low-, medium-, high-, and mixed-SES schools after Kernel matching

	Comparison 1		Comparison 2		Comparison 3		Comparison 4		Comparison 5		Comparison 6	
	Low-SES school (a) versus medium-SES school (b)		Low-SES school (a) versus high-SES school (b)		Low-SES school (a) versus mixed-SES school (b)		Medium-SES school (a) versus high-SES school (b)		Medium-SES school (a) versus mixed-SES school (b)		High-SES school (a) versus mixed-SES school (b)	
	b(d)	*SE*	*b(d)*	*SE*	*b(d)*	*SE*	*b(d)*	*SE*	*b(d)*	*SE*	*b(d)*	*SE*
Intercept[a]	67.33***	0.39	66.22***	0.81	67.75***	0.62	66.88***	0.84	68.56***	0.57	68.79***	0.73
School (b)[b]	0.93	2.07	−1.04	2.44	3.15	2.15	1.39	1.21	2.58*	1.05	0.82	1.53
SES	2.41***	0.32	2.01***	0.39	1.82***	0.40	2.11***	0.48	2.50***	0.40	2.42***	0.42
Intercept*school (2)*SES	0.66	1.22	1.25	1.14	1.37	1.09	0.48	0.52	0.15	0.44	−0.43	0.70
Growth phase 1	11.86***	0.46	12.75***	0.86	11.90***	0.34	11.42***	1.58	11.02***	0.47	10.87***	0.71
Growth phase 2	20.85***	0.64	21.73***	1.37	13.22***	1.02	20.40***	1.12	19.52***	0.94	19.38***	1.20
Growth phase 3	28.60***	0.56	30.69***	1.01	26.73***	0.91	28.35***	1.28	26.11***	0.86	26.29***	0.96
Growth phase 4	34.28***	0.68	36.34***	1.11	31.99***	1.14	34.26***	0.90	30.93***	0.99	31.52***	1.12
Growth phase 5	38.51***	0.54	38.42***	1.27	36.01***	1.01	37.13***	1.15	34.43***	1.14	35.63***	1.18
Growth phase 1*school (b)	2.56	2.59	7.26*	3.33	−0.68	2.90	0.10	1.23	−2.37*	1.16	−2.10	1.90
Growth phase 2*school (b)	2.33	1.86	1.89	3.18	−0.49	2.13	0.86	1.62	−1.78	1.19	−0.97	1.80
Growth phase 3*school (b)	−0.82	1.76	−4.31	2.70	−2.47	2.20	0.70	1.17	−1.50*	0.76	−0.80	1.10
Growth phase 4*school (b)	−1.51	1.17	−5.25**	1.76	−1.80	1.14	0.12	1.13	−0.99	0.72	−0.33	1.17
Growth phase 5*school (b)	−1.59	1.37	4.23*	1.94	−2.59	1.36	−0.70	0.31	−0.17	0.70	0.25	0.82
Growth phase 1*SES	0.14	0.42	−0.21	0.91	1.16**	0.44	−0.36	0.43	0.80*	0.37	1.32**	0.45
Growth phase 2*SES	0.62	0.43	2.29**	0.77	0.73	0.52	0.03	0.62	0.04	0.43	0.67	0.51
Growth phase 3*SES	0.32	0.37	1.17	0.99	0.01	0.49	1.14	0.46	0.04	0.37	0.63	0.44
Growth phase 4*SES	0.58	0.47	2.50	1.46	0.73	0.54	0.79	0.45	0.38	0.46	1.02*	0.42
Growth phase 5*SES	0.65	0.42	2.15	0.52	0.50	0.49	1.30**	0.50	0.57	0.43	0.63	0.33
Growth phase 1*school (b)*SES	0.41	1.11	0.69	1.29	−0.62	1.09	0.95	0.51	−0.28	0.38	−1.62**	0.57
Growth phase 2*school (b)*SES	−0.97	2.21	−2.67	2.57	−1.19	2.11	0.75	0.65	0.65	0.46	−0.60	0.78
Growth phase 3*school (b)*SES	−2.58	1.11	−3.20	1.77	−2.22*	1.10	0.59	0.51	0.54	0.39	−0.29	0.62

Continued

Table 9. (*Continued*)

	Comparison 1		Comparison 2		Comparison 3		Comparison 4		Comparison 5		Comparison 6	
	Low-SES school (a) versus medium-SES school (b)		Low-SES school (a) versus high-SES school (b)		Low-SES school (a) versus mixed-SES school (b)		Medium-SES school (a) versus high-SES school (b)		Medium-SES school (a) versus mixed-SES school (b)		High-SES school (a) versus mixed-SES school (b)	
	b(*d*)	*SE*	*b*(*d*)	*SE*	*b*(*d*)	*SE*	*b*(*d*)	*SE*	*b*(*d*)	*SE*	*b*(*d*)	*SE*
Growth phase 4*school (b)*SES	−0.68	1.85	−2.97	2.59	−1.30	1.81	0.19	0.47	0.52	0.49	0.07	0.51
Growth phase 5*school (b)*SES	−0.56	1.45	−2.14	1.70	−0.63	1.54	0.06	0.52	0.84	0.47	0.52	0.51
School ethnic composition	−0.33	0.67	−0.60	0.78	0.46	0.60	−0.11	0.58	0.54	0.49	−0.25	0.78
Growth phase 1*ethnic composition	−1.23	0.62	−1.01	0.90	−1.20	0.74	0.60	0.73	−0.76	0.39	0.82	0.87
Growth phase 2*ethnic composition	−0.18	1.29	0.21	1.61	−0.99	1.03	−0.44	0.69	−0.63	0.58	0.47	1.17
Growth phase 3*ethnic composition	0.24	1.02	0.78	1.41	−0.60	0.93	0.39	0.69	−0.61	0.66	−0.02	1.23
Growth phase 4*ethnic composition	1.17	1.18	1.26	1.70	−0.06	1.21	0.30	0.71	0.03	0.66	0.12	1.15
Growth phase 5*ethnic composition	0.91	0.98	1.27	1.74	−0.04	1.00	−0.40	0.56	−0.18	0.62	0.63	1.10
School math composition	2.10***	0.48	2.00*	0.86	2.17***	0.48	2.41***	0.42	2.41***	0.36	1.25*	0.50
Growth phase 1*math composition	−1.82*	0.71	−3.44	0.92	−1.71*	0.82	−0.95	0.67	−0.57	0.34	0.18	0.98
Growth phase 2*math composition	−1.73	0.99	−4.70	1.34	−2.40*	0.97	−1.17	0.65	−0.49	0.48	−0.15	1.11
Growth phase 3*math composition	−0.84	0.76	−3.31	1.04	0.91	0.65	−1.22*	0.59	−0.50	0.41	−0.65	1.00
Growth phase 4*math composition	−0.98	0.91	−2.89	1.26	−1.49	0.88	−1.15*	0.56	−0.47	0.47	−0.32	0.88
Growth phase 5*math composition	−1.41*	0.63	−2.66	1.31	−1.30	0.83	−0.32**	0.46	−0.86	0.50	0.28	1.02
School size	0.07	0.44	2.96**	0.88	−0.19	0.69	0.62	0.44	0.08	0.32	1.17	0.73
Growth phase 1*school size	0.47	0.48	−2.82	1.70	1.48*	0.69	−0.92	0.71	0.39	0.32	−0.67	0.96
Growth phase 2*school size	0.39	0.68	−4.91**	1.45	1.36	1.00	−1.04	0.58	0.12	0.43	−0.81	1.29
Growth phase 3*school size	−0.24	0.64	−4.52***	1.17	1.48	0.92	−0.57	0.50	0.32	0.40	0.21	1.06
Growth phase 4*school size	0.56	0.75	−2.38*	1.08	1.25	1.11	−0.12	0.38	0.19	0.40	0.34	0.91
Growth phase 5*school size	−0.51	0.59	−2.93**	0.95	1.67	1.05	−0.43	0.44	0.48	0.45	0.11	1.26

Note. [a]The intercept refers the math score of a student in school (a) of each comparison at the end of Grade 1.
[b]School (b) refers to the value-added of school (b) in each comparison.
$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

that comparable students showed more positive math achievement growth in high-SES as compared to low-SES and mixed-SES schools. Furthermore, in two of the three comparisons, students in mixed-SES schools showed the lowest math development. More specifically, students in both high-SES and medium-SES schools showed more positive math achievement growth in comparison with similar peers in mixed-SES schools. Differential analyses by individual SES further revealed that the negative relationship between mixed-SES schools and math achievement growth appeared to be the strongest for high-SES students.

Our finding that students in high-SES schools showed higher math achievement as compared to students in low-SES schools is in line with findings of other research on this topic (Agirdag *et al.*, 2012; Battistich *et al.*, 1995; Driessen, 2002; Dumay & Dupriez, 2008; Peetsma *et al.*, 2006; Strand, 1997; Van der Slik *et al.*, 2006; Willms, 2010). Different explanations have been advanced for this positive effect of high-SES schools. For example, it has been suggested that school tend to adapt the general difficulty level to the average level of their students. As such, in schools with large numbers of children whose parents have had little schooling, teachers tend to lower the educational level by focusing more on basic skills and less on higher order skills (Peetsma *et al.*, 2006). It has also been found that teachers hold higher expectations of their students with students' potential to achieve in high-SES than in low-SES schools (Rubie-Davies, Flint, & McDonald, 2012; Timmermans, Kuyper, & Werf, 2015). Research has shown that when teachers have high expectations of students, students tend to confirm the expectations, a phenomenon which is known as 'self-fulfilling prophecy' (Rosenthal & Jacobson, 1968). In addition, it has been argued that in low-SES schools, students experience lower peer pressure levels to succeed academically, which might discourage them from working harder (OECD, 2001). Finally, it has been proposed that high-SES schools benefit more from parent support (Opdenakker *et al.*, 2002).

To our knowledge, only two studies have studied the effect of school SES on math achievement growth throughout primary education (Guldemond & Bosker, 2009; Verhaeghe *et al.*, 2011). Both studies concluded that although low-SES schools were associated with lower achievement at the first measurement comparison, no associations could be found with achievement growth. As such, these studies' findings are opposite to the findings of the present study, as we found a positive association between low-SES school composition and achievement at the first measurement occasion but a negative relation between low-SES school composition and achievement growth. There are several reasons for why our findings may be different from these two previous studies. First, these two studies did not control as extensively for selection bias as our study and did not apply a quasi-experimental approach. Second, these studies did not control for the achievement composition of the school. As such, it might be possible that the lower achievement they found to be associated with the first measurement occasion was the sheer result of the differences in background characteristics or the lower achievement level that is often found in low-SES schools, as these two types of school composition are highly correlated (Belfi *et al.*, 2015; Willms, 2006, 2010).

Furthermore, our finding that students showed more achievement growth in more homogenous schools (i.e., high- and medium-SES schools) than in heterogonous schools where there was a mix in terms of student SES, are in accordance with research findings on homogenous versus heterogeneous school achievement compositions (Hong, Corter, Hong, & Pelletier, 2012; Hoxby & Weingarth, 2005; Pinto, 2012). Several causal paths have been proposed through which the beneficial effect of homogeneous grouping could take place. First, children with different achievement levels, but also children with different SES levels, differ in terms of interests, prior experiences, readiness, and learning profiles. While it is generally the goal of teachers to ensure that every student learns effectively, research has shown that class groupings with a wide variety in student types, present teachers with complex and difficult pedagogical dilemma's (Tomlinson *et al.*, 2003). Even when teachers attempt to apply differentiated instruction, this is often made in ways that are limited or ineffective. Teachers seem particularly resistant to adapting or modifying materials, planning lessons for individuals, and changing evaluation procedures (Schumm *et al.*, 1995). Research has further shown that in heterogeneous classes with a mosaic of students, especially the more able students tend to fall short (Hong *et al.*, 2012; Tomlinson *et al.*, 2003). This is because lower ability children, who on average have a lower capacity for self-regulation, typically need more direct guidance and more frequent assistance from the teacher (Hong *et al.*, 2012). Furthermore, it has been indicated that teachers often lack knowledge on how to modify the curriculum for students whose proficiencies extend beyond those prescribed by grade-level curricula (Tomlinson *et al.*, 2003). As SES and achievement are highly related (Belfi *et al.*, 2015), these examples could also offer an explanation for our finding that especially students with higher individual SES levels are disadvantaged by heterogeneous, mixed-SES school settings. This suggests that teachers should invest more equally in low- and high-SES students, as every student deserves the same possibilities to develop as much as he or she can.

Strengths and limitations

We believe that the present study has made a meaningful contribution to the available research on primary school socio-economic composition. To make this contribution, we have used data from a large primary school sample, followed students throughout primary education, matched students attending schools with different socio-economic composition types on a large set of student characteristics, and studied differential effects by student SES.

However, apart from these strengths, some limitations should be noted. A first concern may stem from the propensity score matching method itself. It still cannot be definitely ruled out that difference between the school SES types is not the result of selection bias on unobservables (Becker, Lüdtke, Trautwein, Köller, & Baumert, 2012). Such bias may be caused by unreliability of the measures or by failure to control for unobserved confounding variables. In this case, the samples would not be fully comparable, and the differences observed in the outcome variable might be explained by the remaining heterogeneity of the groups. This is a common problem of methods relying on the control of observed characteristics as regression and propensity score matching and cannot be resolved with the existing data. A different study design, for example a longitudinal study with several measurement occasions before the transition to primary school, would help to address this problem by gauging the degree of change in the development trajectory after the transition to different school socio-economic compositions relative to the students' previous development (Raudenbush, 2001). Nonetheless, the model presented here contains more information to control for selection bias than has been the case in previous research. At the same time, it makes less extreme statistical assumptions and gives more information about the generalizability of the effects than regression analysis.

Second, because school SES was measured categorically, some variation within this variable was not considered. It is possible that small variations within the different categories may have led to small differences in the results. This is an inherent problem of the propensity score matching framework (Crosnoe, 2009). Yet, we believe that the advantages of the propensity score matching approach (i.e., better control for selection bias and less extrapolation of portions of the covariate space where there are no data) outweigh this disadvantage.

Finally, students' achievement development was only studied with respect to mathematical development. To obtain a more detailed picture of how socio-economic school composition affects students' achievement development, other domains, such as language and science, should be studied. These limitations should be addressed in future research.

Conclusions

In sum, we have found a clear effect of school socio-economic composition on student math achievement growth throughout primary education. More specifically, high-SES schools were found to be most beneficial for students' math development and mixed-SES schools the least. The negative effects of mixed-SES schools (as compared to medium- and high-SES schools) were found to be most prominent for high-SES students. Despite some limitations, our study suggests that propensity score matching might be a viable way for studying the effects of school socio-economic composition.

References

Agirdag, O., Van Houtte, M., & Van Avermaet, P. (2012). Why does the ethnic and socioeconomic composition of schools influence math achievement? The role of sense of futility and futility culture. *European Sociological Review*, *28*, 366–378. doi:10.1093/esr/jcq070

Battistich, V., Solomon, D., Kim, D., Watson, M., & Schaps, E. (1995). Schools as communities, poverty levels of student populations, and students' attitudes, motives, and performance: A multilevel analysis. *American Educational Research Journal*, *32*, 627–658. doi:10.3102/00028312032003627

Becker, M., Lüdtke, O., Trautwein, U., Köller, O., & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology*, *104*, 682–699. doi:10.1037/a0027608

Belfi, B., Gielen, S., De Fraine, B., Verschueren, K., & Meredith, C. (2015). School-based social capital: The missing link between schools' socioeconomic composition and collective teacher efficacy. *Teaching and Teacher Education*, *45*, 33–44. doi:10.1016/j.tate.2014.09.001

Belfi, B., Goos, M., Pinxten, M., Verhaeghe, J. P., Gielen, S., De Fraine, B., & Van Damme, J. (2014). Inequality in language achievement growth? An investigation into the impact of pupil socio-ethnic background and school socio-ethnic composition. *British Educational Research Journal*, *40*, 820–846. doi:10.1002/berj.3115

Benner, A. D., & Crosnoe, R. (2011). The racial/ethnic composition of primary schools and young children's academic and socioemotional functioning. *American Educational Research Journal*, *48*, 621–646. doi:10.3102/0002831210384838

Caliendo, M., & Kopenig, S. (2005). *Some practical guidance for the implementation of propensity score matching*. Berlin, DIW German Institute for Economic Research. Discussion Paper 485.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Cortois, L., Van Droogenbroeck, I., Verachtert, P., & Van Damme, J. (2011). *Longitudinaal onderzoek in het basisonderwijs: Toetsen zesde leerjaar (schooljaar 2008–2009). [A longitudinal study in primary education: Achievement tests sixth grade (school year 2008–2009)]*. SSL Report 32.

Crosnoe, R. (2009). Low-income students and the socioeconomic composition of public high schools. *American Sociological Review*, *74*, 709–730. doi:10.1177/000312240907400502

Dors, H. G., Karsten, S., Ledoux, G., & Steen, A. H. M. (1991). *Etnische segregatie in het onderwijs. Beleidsaspecten [Ethnic segregation in Education. Policy aspects]*. Amsterdam: SCO-Kohnstamm Instituut.

Driessen, G. (2002). School composition and achievement in primary education: A large-scale multilevel approach. *Studies in Educational Evaluation*, *28*, 347–368. doi:10.1016/S0191-491X(02)00043-3

Dumay, X., & Dupriez, V. (2008). Does the school composition effect matter? Evidence from Belgian data. *British Journal of Educational Studies*, *56*, 440–477. doi:10.1111/j.1467-8527.2008.00418

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevelhierarchical models* (Vol. 1). Cambridge, UK: Cambridge University Press.

Guldemond, H., & Bosker, R. J. (2009). School effects on students' progress – a dynamic perspective. *School Effectiveness and School Improvement*, *20*, 255–268. doi:10.1080/09243450902883938

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, *15*, 234–249. doi:10.1037/a0019623

Hattie, J. A. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, *37*, 449–481. doi:10.1016/S0883-0355(03)00015-6

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*, 199–236. doi:10.1093/pan/mpl013

Hong, G., Corter, C., Hong, Y., & Pelletier, J. (2012). Differential effects of literacy instruction time and homogeneous ability grouping in kindergarten classrooms who will benefit? Who will suffer? *Educational Evaluation and Policy Analysis*, *34*, 69–88. doi:10.3102/0162373711424206

Hoxby, C. M., & Weingarth, G. (2005). *Taking race out of the equation: School reassignment and the structure of peer effects*. Working paper.

Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. Hoyle (Ed.), *Structural equation modeling: Issues, concepts, and applications* (pp. 76–99). Newbury Park, CA: Sage.

Imai, K., King, G., & Stuart, A. (2008). Misunderstandings among experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *171*, 481–502. doi:10.1111/j.1467-985X.2007.00527.x

Jacobs, D., Rea, A., & Teney, C. (2009). *De sociale lift blijft steken. De prestaties van allochtone leerlingen in de Vlaamse Gemeenschap en de Franse Gemeenschap [The social lift stops working. The performance of immigrant students in the Flemish community and the French community]*. Brussels, Belgium: Koning Boudewijnstichting.

Langenskiöld, S. (2005). *Peer influence on smoking: Causation or correlation?* (Unpublished doctoral dissertation). Stockholm School of Economics, Stockholm, Sweden.

Leuven, E., & Sianesi, B. (2003). *PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing*. EconPapers. Retrieved from http://econpapers.repec.org/software/bocbocode/S432001.htm

Maes, F., Ghesquiere, P., Onghena, P., & Van Damme, J. (2002). *Longitudinaal onderzoek in het basisonderwijs. Van doelstellingen tot onderzoeksopzet* [A longitudinal study in primary education. Research objectives and planning]. Leuven, Belgium: Steunpunt Loopbanen doorheen Onderwijs naar Arbeidsmarkt.

Manski, C. F. (2000). Economic analysis of social interactions. *The Journal of Economic Perspectives*, *14*, 115–136. doi:10.3386/w7580

OECD (2001). *Knowledge and skills for life: First results from the OECD programme for International Student Assessment (PISA), 2000 (no. 92-64-19671-4)*. Paris, France: Author.

OECD (2010). *PISA 2009 results: Overcoming social background – Equity in learning opportunities and outcomes* (Vol. II). Paris, France: Author.

Opdenakker, M. C., Van Damme, J., De Fraine, B., Van Landeghem, G., & Onghena, P. (2002). The effect of schools and classes on mathematics achievement. *School Effectiveness and School Improvement*, *13*, 399–427. doi:10.1076/sesi.13.4.399.10283

Peetsma, T., Van der Veen, I., Koopman, P., & Van Schooten, E. (2006). Class composition influences on students' cognitive development. *School Effectiveness and School Improvement*, *17*, 275–302. doi:10.1080/13803610500480114

Pinto, Y. (2012). The efficacy of homogeneous groups in enhancing individual learning. *Journal of Education and Practice*, *3*, 25–38.

Pinxten, M., De Fraine, B., Van Damme, J., & D'Haenens, E. (2010). Causal ordering of academic self-concept and achievement: Effects of type of achievement measure. *British Journal of Educational Psychology*, *80*, 689–709. doi:10.1348/000709910X493071

Raudenbush, S. W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology*, *52*, 501–525. doi:10.1146/annurev.psych.52.1.501

Retelsdorf, J., Becker, M., Köller, O., & Möller, J. (2012). Reading development in a tracked school system: A longitudinal study over 3 years using propensity score matching. *British Journal of Educational Psychology*, *82*, 647–671. doi:10.1111/j.2044-8279.2011.02051

Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55. doi:10.1093/biomet/70.1.41

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart, & Winston.

Royston, P. (2004). Multiple imputation of missing values. *The Stata Journal*, *4*, 227–241.

Rubie-Davies, C. M., Flint, A., & McDonald, L. G. (2012). Teacher beliefs, teacher characteristics, and school contextual factors: What are the relationships? *British Journal of Educational Psychology*, *82*, 270–288. doi:10.1111/j.2044-8279.2011.02025.x

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177. doi:10.1037//1082-989X.7.2.147

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods*, *13*, 279–313. doi:10.1037/a0014268

Schumm, J., Vaughn, S., Haager, D., McDowell, J., Rothlein, L., & Saumeil, L. (1995). General education teacher planning: What can students with learning disabilities expect? *Exceptional Children*, *61*, 335–352. doi:10.1177/001440299506100403

Snijders, T. A., & Bosker, R. J. (2012). *Multilevel analysis. Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.

StataCorp. (2015). *Stata 14 base reference manual*. College Station, TX: Author.

Strand, S. (1997). Student progress during Key Stage 1: A value added analysis of school effects. *British Educational Research Journal*, *23*, 471–487. doi:10.1080/0141192970230406

Timmermans, A. C., Kuyper, H., & Werf, G. (2015). Accurate, inaccurate, or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education? *British Journal of Educational Psychology*, *85*, 459–478. doi:10.1111/bjep.12087

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., . . . Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, *27*, 119–145. doi:10.1177/016235320302700203

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*, 1–67.

Van der Slik, F. W. P., Driessen, G. W. J. M., & De Bot, K. L. J. (2006). Ethnic and socioeconomic class composition and language proficiency: A longitudinal multilevel examination in Dutch primary schools. *European Sociological Review*, *22*, 293–308. doi:10.1093/esr/jci058

Van Maele, D., & Van Houtte, M. (2011). The quality of school life: Teacher-student trust relationships and the organizational school context. *Social Indicators Research*, *100*, 85–100. doi:10.1007/s11205-010-9605-8

Verhaeghe, J. P. (2010). *Technische fiche voor IRT-analyse wiskunde basisonderwijs [Technical report IRT analysis math primary education]*. Leuven, Belgium: Steunpunt Loopbanen doorheen Onderwijs naar Arbeidsmarkt.

Verhaeghe, J., Van Damme, J., & Knipprath, H. (2011). *Value added of primary schools with high proportions of minority students: A longitudinal study*. Paper presented at ICSEI 24, Limassol, Cyprus.

Willms, J. D. (2006). *Learning divides: Ten policy questions about the performance and equity of schools and schooling systems*. Montreal, Canada: UNESCO Institute for Statistics.

Willms, J. D. (2010). School composition and contextual effects on student outcomes. *Teachers College Record*, *112*, 1008–1037.

Received 12 May 2015; revised version received 21 April 2016